CLAIMS

What is claimed is:
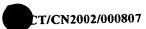
5   1.   A method comprising:
          receiving user input at a client device;
          interpreting the user input to identify a selection of at least one of a plurality of web interaction modes;
          producing a corresponding client request based in part on the user input and the web
10   interaction mode; and
          sending the client request to a server via a network.

     2.   The method as claimed in Claim 1 further including:
          identifying a focused display element, the client request based in part on the identified
15   focused display element.

     3.   The method as claimed in Claim 2 further including:
          sending an identifier of the identified focused display element to the server.

20   4.   The method as claimed in Claim 2 wherein the focused display element is a
     hyperlink.

     5.   The method as claimed in Claim 2 wherein the focused display element is a
     field in a form.
25
     6.   The method as claimed in Claim 1 further including:
          extracting speech features from the user input, the client request based in part on the
     extracted speech features.

30   7.   The method as claimed in Claim 6 further including:
          sending the extracted speech features to the server.

     8.   The method as claimed in Claim 1 further including:
          sending a session message to the server to initialize a connection with the server.

9.   The method as claimed in Claim 8 wherein the session message includes an IP address of the client device, a device type of the client device, a voice character of the user, a language that the user speaks, and a default recognition accuracy that the client device requests.

5

10.   The method as claimed in Claim 1 further including:

sending a transmission message to the server to exchange transmission parameters with the server.

10   11.   The method as claimed in Claim 1 further including:

sending an OnFocus message to the server when a talk button is activated to notify the identifier of a focused display element, and the URL of current page.

12.   The method as claimed in Claim 11 further including:

15   sending extracted speech features to the server.

13.   The method as claimed in Claim 1 further including:

the cases to occur Unfocus message and tasks when Unfocus message occurs.

20   14.   The method as claimed in Claim 1 further including:

sending an exit message to the server to terminate a session with the server.

15.   The method as claimed in Claim 1 wherein a multi-modal markup language is used.

25

16.   A method comprising:

receiving at a server a client request from a client device via a network;

interpreting the client request to identify a selection of at least one of a plurality of web interaction modes, at least one web interaction mode being a speech interaction mode;

30   and

if the speech interaction mode is selected,

receiving an identifier of a focused display element,

building a correct grammar for speech recognition based on the

focused display element,

performing speech recognition, and

· performing specific tasks according to the result of the speech recognition.

17. The method as claimed in Claim 16 wherein the focused display element is a hyperlink.

18. The method as claimed in Claim 16 wherein the focused display element is a field in a form.

19. The method as claimed in Claim 16 further including:

sending a match event to the client device via the network.

20. The method as claimed in Claim 16 further including:

sending a nomatch event to the client device via the network.

21. The method as claimed in Claim 16 further including:

receiving a transmission message from the client device for the exchange of transmission parameters with the client device.

22. A client device comprising:

a user input receiver;

an interpreter to identify a selection of at least one of a plurality of web interaction modes from user input received by the user input receiver, at least one web interaction mode being a speech interaction mode;

a client request generator to generate a client request based in part on the user input and the web interaction mode, and to send the client request to a server via a network.

23. The client device as claimed in Claim 22 wherein the client request generator also identifies a focused display element, the client request based in part on the identified focused display element.

24.   The client device as claimed in Claim 22 wherein the client request generator also sends an identifier of the identified focused display element to the server.

25.   The client device as claimed in Claim 23 further including a web interaction mode interpreter.

26.   A server apparatus comprising:
    a client request receiver to receive a client request from a client device via a network;
    an interpreter to identify a selection of at least one of a plurality of web interaction modes from the client request received by the client request receiver, at least one web interaction mode being a speech interaction mode;
    a speech processor to process speech received in the client request if the speech interaction mode is selected, the speech processor using an identifier of a focused display element, and building a correct grammar for speech recognition based on the focused display element, the speech processor performing speech recognition, and performing specific tasks according to the result of the speech recognition.

27.   The server apparatus as claimed in Claim 26 wherein the focused display element is a hyperlink.

28.   The server apparatus as claimed in Claim 26 wherein the focused display element is a field in a form.

29.   The server apparatus as claimed in Claim 26 further including a web interaction mode interpreter.

30.   A multi-modal network interaction system comprising:
    a client device having a user input receiver, an client interpreter to identify a selection of at least one of a plurality of web interaction modes from user input received by the user input receiver, at least one web interaction mode being a speech interaction mode, and a client request generator to generate a client request based in part on the user input and the web interaction mode, and to send the client request to a server via a network; and

a server having a client request receiver to receive the client request from the client device via the network, a server interpreter to identify a selection of at least one of a plurality of web interaction modes from the client request received by the client request receiver, at least one web interaction mode being a speech interaction mode, and

a speech processor to process speech received in the client request if the speech interaction mode is selected, the speech processor using an identifier of a focused display element, and building a correct grammar for speech recognition based on the focused display element, the speech processor performing speech recognition, and performing specific tasks according to the result of the speech recognition.

31.   The system claimed in Claim 30 wherein the client request generator also identifies a focused display element, the client request based in part on the identified focused display element.

32.   The system as claimed in Claim 31 wherein the client request generator also sends an identifier of the identified focused display element to the server.

33.   The system as claimed in Claim 30 wherein the focused display element is a hyperlink.

34.   The system as claimed in Claim 30 wherein the focused display element is a field in a form.

35.   A machine-readable medium having instructions which when executed cause a machine to perform the method comprising:

receiving user input at a client device;

interpreting the user input to identify a selection of at least one of a plurality of web interaction modes, at least one web interaction mode being a speech interaction mode;

producing a corresponding client request based in part on the user input and the web interaction mode; and

sending the client request to a server via a network.

36.   The machine-readable medium as claimed in Claim 35 further including instructions for:

identifying a focused display element, the client request based in part on the identified focused display element.

5

37.   The machine-readable medium as claimed in Claim 36 further including instructions for:

sending an identifier of the identified focused display element to the server.

10   38.   The machine-readable medium as claimed in Claim 35 wherein the focused display element is a hyperlink.

39.   The machine-readable medium as claimed in Claim 35 wherein the focused display element is a field in a form.

15

40.   A method comprising:

a set of markup language has been defined for applications quickly building over web by multi-modal interaction.

20   41.   A method as claimed in Claim 40 further including:

a conformance definition for the event handling of multi-modal markup language.

42.   A method claimed in Claim 40 further including:

for synchronization, two element's blocks are defined. One is sent to client and the

25   other is kept in server.